Appendix A
Data Sets

APPENDIX A
TABLE OF CONTENTS

Appendix A Comments

Appendix A contains the data for each WAG 9 site that was retained from the preliminary screening for the WAG 9 Remedial Investigation with the Baseline Risk assessment. The soils data from the seven retained sites has been divided into 16 subunits based on the distinct boundaries and potential receptor exposure pathway s. The summary of the subunit screening is shown in Section 3 of this document. The water data was only collected for three retained sites at WAG 9. The samples for the water data was collected for: ANL-01A-MCTBD-Water, ANL-01-Industrial Waste Pond-Water, and for ANL-35-North Ditch-Water.

The data is presented in each of the following 16 soils and three water subsections of Appendix A. To aid in the review and understanding of the data numerous methods of presenting the data were used. Not all methods of presenting the data were used for all retained sites. These presentation methods may include:

- Plan maps showing the location of the combined sampling data
- Color intensity profile maps of the 1994 data
- Scatter plots of all data
- Screening tables for metals, radionuclides, organics which show the compound, maximum concentration, INEL background, detect samples above background, and detect samples
- Statistics for retained contaminants for the groundwater (all data) exposure pathway
- Statistics for the retained contaminants for the residential (0-10') exposure pathway
- Statistics for the retained contaminants for the surface (0.-5') exposure pathway
- and finally the combined data set that shows contaminant, concentration, data qualifiers, sample location, and date sample was collected.

The plan maps show the relative location of all samples collected from the past sampling activities in each retained site. The legend shows the symbol used to differentiate the samples for the different sampling years and events.

The color intensity profile maps present a visualization of the results of the samples collected in the 1994 ANL-W study. Only the data from the ANL-W 1994 study was used to produce these profile maps because it contained the largest data set, and to eliminate the problems of combining data sets with different detection limits and analytical laboratories. The color intensity profile maps show graphically the highest concentrations and ultimately "hot spots" along the ditch units. This aids the reviewer in understanding the location of the "hot spots", the extent of the "hot spots", and source of the "hot spots". The same data that is shown in the simple color profile maps, is also presented in pages and pages of data at the end of each retained subunit. The jagged bottom of the maps shows the depth of the alluvial material on top of the basalt. The methodology used to make the color profile maps is presented at the end of this Appendix A table of contents section.

Scatter plots for the retained sites aids the reviewer in understanding the combined data set for each retained site. The data presented in the scatter plots is similar to what is presented in the color intensity profile maps with the exception that the scatter plots shows all the data for the retained site. The scatter plots show how the concentrations of the contaminants vary along the length of each site. Also, the scatter plots show how the differences in the detection limits for each of the previous sampling events.

The screening tables for the metals, radionuclides, and organics present the same data that is shown in Section 3 for each retained site. The screening tables show the maximum sample, INEL background, number of samples that exceeded INEL background, and number of samples detected. A contaminant was retained if it exceeded background or contaminant was detected and no INEL background value existed.

The inorganic and organic contaminants that were retained were then used to determine separate the data into the exposure pathways based on the depth of the samples. For the groundwater pathway, all of the

data for the site was used to determine the 95% upper one-tailed Upper Confidence Level of the mean. For the residential exposure pathway, only the data collected between 0-10' of the surface was used to determine the calculated and summarized for the 95% upper one-tailed Upper Confidence Level of the mean. Likewise for the surface exposure pathway on the data collected between 0-0.5' of the surface was used to determine the 95% upper one-tailed Upper Confidence Level of the mean. For the radionuclide contaminants the statistics were calculated in a similar manor with the exception that the surface exposure pathway the depth used was 0-4' verses the 0-0.5' for the inorganics and organics. The smallest value between the calculated 95% upper one-tailed Upper Confidence Level of the mean value and the maximum value for each contaminant was used in the risk assessment.

For each retained unit, the combined data set is the last thing presented in each subunit. The combined data set shows the complete set of data for each subunit. Thie combined data set shows the sample number, sample identification, analysis, compound, concentration, radioactive uncertainty, data qualifiers, matrix, type location, location, depth range, and sample date. This data was used in the screening of each subunit and to calculate the statistics and determine the 95% upper one-tailed Upper Confidence Level of the mean values to be used in the risk assessment. Also, to aid in the understanding of what is actually happening in each unit, the data was used to produce the scatter plots and the color intensity profile maps.

The surface water data that was collected for the three sites (ANL-01A-MCTBD-Water, ANL-01-Industrial Waste Pond-Water, and for ANL-35-North Ditch-Water) is presented as a data dump only. The data was not screened against INEL background since there is not INEL background for surface water. However, the hydrogeology section qualitatively discusses comparisons of the data to the Maximum Contaminant Levels (MCL's) which have been established by EPA for drinking water.

ANL-W Vertical Profile Maps

A road map for creating vertical profile maps using ARCINFO

Introduction

This document details the steps taken to build vertical profile maps of 1994 arsenic, beryllium, copper, lead, mercury, nickel, vanadium, and zinc concentrations for selected sites within the Argonne National Laboratory West facility. Section one describes where the data sets used in the project originated. Section two details the steps taken to convert the data from planner coordinates to vertical depths. And finally, step three describes how the data was griddec using the ARCINFO Grid Module.

Section 1 - Data sets used

Two different sets of data were required for this project; quantity data and positional data. The quantity data was pulled from an Unqualified ERIS database maintained by Environmental Restoration (ER). This is an Oracle database used to house sample results data from ER sampling projects. The files that were pulled out the ERIS database contained sample numbers and concentrations for the elements listed above. The positional data was produced at ANL-Chicago using ARC/INFO by digitizing each sample, borehole, location from paper maps. The resulting ARC/INFO electronic maps, coverages, contained positions on the ground where each borehole was drilled, and the accompanying INFO database tables contained the sample numbers and associated depths for each sample that was taken. Because each borehole contained many samples, the database files had a one to many relationship with the actual location on the ground. In order to get the data back to a one to one relationship, one sample location for one sample result, positions for each sample had to be calculated.

Section 2 - Data conversion

The point coverage provided by ANL-Chicago contained only the X and Y coordinates for the point on the surface of the ground, borehole, where each sample was taken. In order to produce a vertical view of the data, new coordinates had to be calculated for each sample. Before the calculation could be made however, the X and Y coordinates had to be extracted from the INFO file for the selected sample locations. The coordinates for each area, MCTBD, INTERCEPTOR CANAL, etc., had to be extracted from the INFO table and written to a file along with the SAMPLE_NUM. These files were then turned into spread sheets where the coordinates could be calculated.

The conversion was accomplished by first calculating a new X coordinate using the follow formula:

$$\text{new } X = ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{.5}$$

The new X coordinates, while distorting positions which do not follow a line, maintain the distance between points.

Next, a new Y coordinate was calculated by multiplying the depth for each sample by some negative number which would make the Z axis (the new Y) proportional to the X axis on the final map page. In other words, the depth had to be exaggerated for sample areas which covered long distances in order to see the concentration gradients on the map.

After the conversions for the coordinates were made, (the old X and Y's replaced with the newly calculated positions) the spread sheet files were turned back into an ARCINFO coverages which showed the sample points vertically. The new coverage was then related to the existing sample results database via the key SAMPLE_NUM.

Section 3 - Gridding the data

The ARCINFO GRID module was used to create colored grids which represent the concentration gradients between sample locations. The darker the color (red) the higher the concentration and the lighter the color (yellow) the lower the concentration. The points where converted to a grid (array) of

concentration values using a kriging algorithm in GRID. This algorithm assigned values to each grid cell in the following manner:
The spatial variation was quantified by the semi-variogram. The semi-variogram was estimated by the sample semi-variogram which was computed from the input point data set. The value of the sample semi-variogram for a separation distance of h (referred to as the lag) is the average squared difference in z value between pairs of input sample points separated by h. The sample semi-variogram was calculated from the sample data with the equation:

$$y \times (h) = 1/2n \sum_{t=1}^{R} \{Z(xi) - (xi + H)\}^2$$

where n is the number of pairs of sample points separated by distance h. The semi-variogram was modeled by fitting a theoretical function to the sample semi-variogram.

For the data points used in this project, the modeling method which proved to be the best fit with respect to the semi-variogram was Spherical. The formula for the Spherical model is:

$$y(h) = co + (3h/2a - 1/2(h/a)^3)$$

The kriging process created grids (arrays) of data with a cell size of .5 meters and a concentration value in each. After each .5 meter cell had a value assigned to it, a color for each value was chosen. A shadeset of 255 shades from yellow to red was produced. Next, a linear stretch of the concentration values over the whole 255 colors was performed resulting in a colored grid based on concentration. The higher the concentration, relative for each grid, the darker the red and the lower the concentration the lighter the yellow.